UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/706,102	11/12/2003	Christopher J. Brockett	M61.12-0549	7706

27366 7590 08/22/2007

WESTMAN CHAMPLIN (MICROSOFT CORPORATION)
SUITE 1400
900 SECOND AVENUE SOUTH
MINNEAPOLIS, MN 55402-3319

EXAMINER	
RIDER, JUSTIN W	

ART UNIT	PAPER NUMBER
2626	

MAIL DATE	DELIVERY MODE
08/22/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/706,102	BROCKETT ET AL.	
	Examiner	Art Unit	
	Justin W. Rider	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *12 November 2003*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-30* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-30* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *12 November 2003* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

1. This action is responsive to communications: Application filed 12 November 2003. Claims 1-30 are pending.

Information Disclosure Statement

2. The information disclosure statement(s) (IDS) submitted on 30May 2006 is in compliance with the provisions of 37 CFR 1.97. Accordingly, the examiner has considered the information disclosure statement(s).

Claim Objections

3. Claim 9 is objected to because of the following informalities: Claim 9 is missing a period. Appropriate correction is required.

Claim Rejections - 35 USC § 102

4. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

5. Claims 1-2, 6-17 and 20-24 are rejected under 35 U.S.C. 102(b) as being anticipated by **Shinyama et al. 'Automatic Paraphrase Acquisition from News Articles', Proceedings of Human Language Technology Conference; June 2002** referred to as **Shinyama** hereinafter.

Claim 1: **Shinyama** discloses a method of training a paraphrase system, comprising:

i. receiving a cluster of related texts (p. 2 of 6, section 3.1, *'we find articles of a certain domain from two newspapers.'*);

ii; selecting a set of text segments from the cluster (p. 2 of 6, section 3.1, *'We use an existing IR system to obtain articles from a given class of events,'* [emphasis supplied]); and

iii. using textual alignment to identify paraphrase relationships between text in the text segments in the set (p. 2 of 6, section 3.1, *'In this stage we use a TF/IDF based method...'*).

Claim 2: **Shinyama** discloses a method as per claim 1 above, comprising:

i. using statistical textual alignment to align words in the text segments in the set (p. 3 of 6, section 3.2, step 3, *'We mark all NEs using an statistical NE tagging system [7].'*); and

ii. identifying the paraphrase relationships based on the aligned words (p. 3 of 6, section 3.2, step 4, *'Now we can get paraphrases. First we take pairs of similar sentences...'* e.g. from step 3, *'here, $POST_1$ slot is filled with the actual NE "President")*).

Claim 6: **Shinyama** discloses a method as per claim 1 above, further comprising calculating an alignment model based on the paraphrase relationships identified. It is inherent from within the disclosure of **Shinyama** that an alignment model is calculated due to the fact that section 3 describes the building of a model from aligning/paraphrasing texts and sections 4-6 teach the practical implementation of the training portion on a real application.

Claim 7: **Shinyama** discloses a method as per claim 6 above, further comprising:

i. receiving an input text (page 4 of 6, section 4, *'We used one year of two Japanese newspapers (Mainichi and Nikkei) in this experiment.'*); and

ii. generating a paraphrase of the input text based on the alignment model (page 4 of 6, section 4, *'we ran the paraphrase acquisition system [alignment model] for each pair of articles and finally got total 136 pairs of paraphrases (a link between two IE patterns).'*).

Claim 8: **Shinyama** discloses a method as per claim 1 above, wherein selecting a set of text segments [articles] comprises selecting text segments for the set based on a number of shared words in the text segments (In Figure 2 and section 3.2, **Shinyama** discloses pattern matching for the determination of similar articles. This matching is based on word similarity scoring algorithms and so therefore are directly related to a shared number of words.).

Claim 9: **Shinyama** discloses a method as per claim 1 above, wherein prior to receiving a cluster, identifying the cluster of related texts (p. 4 of 6, section 4, e.g. arrest events and personnel affairs).

Claim 10: **Shinyama** discloses a method as per claim 9 above, wherein identifying a cluster comprises:

i. accessing a plurality of documents (p. 4 of 6, section 4, *'We used one year of two Japanese newspapers,'*); and

ii. identifying documents written by different authors about a common subject, as clusters of related documents (p. 4 of 6, section 4, *'First we obtained the most relevant 300 articles from Mainichi newspaper (total of 11 1373 articles) for two domains, arrest events and personnel affairs,'*).

Claim 11: **Shinyama** discloses a method as per claim 10 above, wherein selecting a text segment set comprises grouping desired text segments of the related documents in each cluster into a set of related text segments [articles] (p. 4 of 6, section 4).

Claim 12: **Shinyama** discloses a method as per claim 11 above, wherein identifying documents comprises identifying documents written within a predetermined time of one another (p. 4 of 6, section 4, documents were used that were within a one year time span.).

Claim 13: **Shinyama** discloses a method as per claim 11 above, wherein accessing a plurality of documents comprises accessing a plurality of different news articles written about a common event (page 4 of 6, section 4, *'We used one year of two Japanese newspapers (Mainichi and Nikkei.'*).

Claim 14: **Shinyama** discloses a method as per claim 13 above, wherein accessing a plurality different news articles comprises accessing a plurality of different news articles written by different news agencies (p. 4 of 6, section 4, *'First we obtained the most relevant 300 articles from Mainichi newspaper (total of 11 1373 articles) for two domains, arrest events and personnel affairs...Next we find the corresponding articles of Nikkei newspaper from 181086 articles (See Table 3).'*).

Claim 15: **Shinyama** discloses a method as per claim 14 above, wherein grouping desired text segments comprises grouping a first predetermined number of sentences of each news article in each cluster into the set of related text segments (p. 4 of 6, section 4, *'After dropping the patterns which appear only once, we got 725 patterns and 157 patterns respectively.'* (Recall that patterns represent sentences from articles that have a similarity score above a certain threshold.)).

Claim 16: **Shinyama** discloses a method as per claim 15 above, wherein selecting a set of text segments comprises pairing each sentence in a given set of related text segments with each

other sentence in the given set (p. 4 of 6, section 4, *'and finally got total 136 pairs of paraphrases (a link between two IE patterns).'*).

Claim 17: **Shinyama**, in view of **Gibson** a paraphrase processing system, comprising a textual alignment component (p. 2 of 6, section 3.1, *'In this stage we use a TF/IDF based method...'*) configured to receive a set of text segments (p. 2 of 6, section 3.1, *'we find articles of a certain domain from two newspapers.'*) and identify paraphrase relationships between words in the set of text segments based on alignment of the words (p. 3 of 6, section 3.2, step 4, *'Now we can get paraphrases. First we take pairs of similar sentences...'* e.g. from step 3, *'here, $POST_1$ slot is filled with the actual NE "President"'*).

Claim 20: **Shinyama** discloses a method as per claim 17 above, further comprising a clustering component configured to access a plurality of documents and cluster the documents based on a subject matter of the documents [articles] (p. 4 of 6, section 4, *'First we obtained the most relevant 300 articles from Mainichi newspaper (total of 11 1373 articles) for two domains, arrest events and personnel affairs,'*).

Claim 21: **Shinyama** discloses a method as per claim 20 above, wherein the clustering component is configured to cluster documents written about a same subject (p. 4 of 6, section 4, *'First we obtained the most relevant 300 articles from Mainichi newspaper (total of 11 1373 articles) for two domains, arrest events and personnel affairs,'*).

Claim 22: **Shinyama** discloses a method as per claim 20 above, wherein the clustering component is configured to extract predetermined text segments from clustered documents to form the set of text segments (p. 4 of 6, section 4, *'We got 294 pairs of articles m arrest events, and 289 pairs of articles in personnel affairs.'*).

Claim 23: **Shinyama** discloses a method as per claim 22 above, further comprising a pairing component configured to identify a plurality of pairs of text segments based on the set of text segments (p. 4 of 6, section 4, *'After dropping the patterns which appear only once, we got 725 patterns and 157 patterns respectively;' 'and finally got total 136 pairs of paraphrases (a link between two IE patterns).'*).

Claim 24: **Shinyama** discloses a method as per claim 23 above, wherein the pairing component is configured to identify the plurality of pairs of text segments by pairing each text segment in a given set of text segments with each other text segment in the given set of text segments (p. 4 of 6, section 4, *'and finally got total 136 pairs of paraphrases (a link between two IE patterns).'*).

Claim Rejections - 35 USC § 103

6. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

7. Claims 3 and 25-30 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Shinyama**.

Claim 3: **Shinyama** discloses a method as per claim 2 above, however **Shinyama** only implicitly teaches the same limitation as claim 2 above merely applied to multi-word phrases. It would have been obvious to one having ordinary skill in the art at the time of invention that this step would be common knowledge in the art, if not common sense to apply the single-word

method to a multi-word application. This is desirable at least for the efficient processing of common multi-word terms (e.g. New York, status quo, prima facie, etc...). If an algorithm can compare two single-word phrases then the ability to compare multi-word phrases would certainly be available with predictable results comparable to that of the single-word variation. Therefore, claim 3 is rejected under the same rationale as in claim 2 above for being of similar scope and content.

Claim 25: **Shinyama** discloses a method as per claim 20 above, however failing to disclose a data store storing the plurality of documents. The examiner is taking Official Notice that it would have been obvious to one having ordinary skill in the art at the time of invention to include a data store for storing documents because providing a system or apparatus with a data store for storing documents is a feature of common knowledge in the computing and signal processing arts. This would be a desirable component because it would allow storage of a large amount of accessible data in which to use to produce documents. The MPEP states that 'the rationale may be expressly or impliedly contained in the prior art or it may be reasoned from knowledge generally available to one of ordinary skill in the art, established scientific principles, or legal precedent established by prior case law.' *See MPEP § 2144.*

Claim 26: **Shinyama** discloses a system as per claim 25 above, wherein the documents consist of a plurality of different news articles written by different news agencies about a common event (page 4 of 6, section 4, *'We used one year of two Japanese newspapers (Mainichi and Nikkei.'*).

Claim 27: **Shinyama** discloses a system as per claim 26 above, wherein the clustering component is configured to cluster the news articles based on a time at which the news articles were written (p. 4 of 6, section 4, documents were used that were within a one year time span.).

Claim 28: Claim 28 is similar in scope and content to that of claim 25 above and so therefore is rejected under the same rationale.

Claim 29: **Shinyama** discloses a system as per claim 17 above, further comprising a paraphrase generator, receiving a textual input and generating a paraphrase of the textual input based on the paraphrase relationships (p. 4 of 6, section 4, *'and finally got total 136 pairs of paraphrases (a link between two IE patterns).'*).

Claim 30: Claim 30 is similar in scope and content to that of claims 17 and 29 and so therefore is rejected under the same rationale.

8. Claims 4-5 and 18-19 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Shinyama** in view of **Gibson et al. (US 2003/0033279 A1)** referred to as **Gibson** hereinafter.

Claim 4: **Shinyama** discloses a method as per claim 1 above further disclosing text alignment in order to identify paraphrase relationships, however failing to, but **Gibson** does distinctly disclose the use of heuristic techniques to perform textual alignment (paragraph [0065], *'The key to the BLAST heuristic is that a statistically significant alignment is likely to contain a high scoring pair (HSP) of aligned words.'*).

Therefore, it would have been obvious to one having ordinary skill in the art at the time of invention to include the teachings of **Gibson** in the methods of **Shinyama** because it provides a highly efficient string searching and alignment method that is easily conducted on any suitable

hardware and that provides an extremely high percentage of high scoring word pairs (p. 4-5, summary of invention).

Claim 5: **Shinyama**, in view of **Gibson** discloses a method as per claim 4 above, however **Shinyama**, in view of **Gibson** only implicitly teaches the same limitation as claim 4 above merely applied to multi-word phrases. It would have been obvious to one having ordinary skill in the art at the time of invention that this step would be common knowledge in the art, if not common sense to apply the single-word method to a multi-word application. If an algorithm can compare two single-word phrases then the ability to compare multi-word phrases would certainly be available with predictable results comparable to that of the single-word variation. Therefore, claim 5 is rejected under the same rationale as in claim 4 above for being of similar scope and content.

Claim 18: **Shinyama**, in view of **Gibson** discloses a method as per claim 17 above, wherein the textual alignment component is configured to generate an alignment model based on statistical (p. 3 of 6, section 3.2, step 3, *'We mark all NEs using an statistical NE tagging system [7].'*) or heuristic alignment of the words (paragraph [0065], *'The key to the BLAST heuristic is that a statistically significant alignment is likely to contain a high scoring pair (HSP) of aligned words.'*).

Therefore, it would have been obvious to one having ordinary skill in the art at the time of invention to include the teachings of **Gibson** in the methods of **Shinyama** because it provides a highly efficient string searching and alignment method that is easily conducted on any suitable hardware and that provides an extremely high percentage of high scoring word pairs (p. 4-5, summary of invention).

Claim 19: **Shinyama**, in view of **Gibson** discloses a method as per claim 18 above, however **Shinyama** only implicitly teaches the same limitation as claims 2 and 4 above merely applied to multi-word phrases. It would have been obvious to one having ordinary skill in the art at the time of invention that this step would be common knowledge in the art, if not common sense to apply the single-word method to a multi-word application. This is desirable at least for the efficient processing of common multi-word terms (e.g. New York, status quo, prima facie, etc...). If an algorithm can compare two single-word phrases then the ability to compare multi-word phrases would certainly be available with predictable results comparable to that of the single-word variation. Therefore, claim 19 is rejected under the same rationale as in claims 2 and 4 above for being of similar scope and content.

Conclusion

9. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Justin W. Rider whose telephone number is (571) 270-1068. The examiner can normally be reached on Monday - Friday 7:30AM - 5:00PM.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, David R. Hudspeth can be reached on (571) 272-7843. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

J.W.R.
07 August 2007

DAVID HUDSPETH
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2